Electricity Costing and Pricing Study

Volume V

Theoretical Foundations of Marginal Cost Pricing

October, 1976

Table of Contents

I. INTRODUCTION

The purpose of this volume is to discuss the theoretical foundations of the recommended cost-pricing philosophy. It presents the thesis that the level of a utility's output is optimal or efficient when price equals the marginal cost of production.

In its earliest form, this theory was stated as a theory of output control: namely, that output should be increased or decreased until price equals marginal cost.

The theory of using marginal-cost pricing to control output is an obvious implication of marginal analysis. This theory says that price, the best available measure of the social gain from marginal output, should equal marginal cost, the best available measure of the social real cost of marginal output. The basic reason is simply that real benefit should be equal to real cost at the margin because this maximizes welfare.

Marginal cost pricing simply means that changes in a customer's total bill resulting from his decision to increase or reduce his use of electricity ought to reflect the additional costs the utility incurs or savings it realizes. That is, rates should track the increased or reduced costs of producing and delivering electricity to the customer that result from his decisions about using it.

The argument that marginal-cost pricing will maximize the well-being of society is founded on the following three propositions:

1. that the input resources are scarce;
2. that the demand for the commodity is sensitive to price;
3. that the customer is the most fitting judge of the mix of goods and service he desires.

Where any one of the conditions does not apply, marginal-cost pricing may not be appropriate.

Marginal costs are based on the system-expansion plan of the planning-period and the cost estimates of the system planner applicable to the facilities to be installed: that is, the marginal capacity costs and marginal energy costs of those least-cost facilities. They are the rate-year present value of cost estimates of the total resources the utility requires from society to generate and distribute electricity for its customers.

While inflation is not allowed for when computing marginal costs, changes in relative prices are taken into account as far as they affect the system planner's least-cost mix of generation and transmission facilities. Insofar as the cost of producing energy varies for different periods of the day and year, the methodology will reflect these variations. Finally, the object of selling at marginal cost is to provide a valid guide to help users in their choice, so that the least-cost solution for them is also the least-cost solution for society.

There are two ways to elaborate on this brief statement of the case for marginal-cost-price output control. First, one could try to state and demonstrate all the assumptions and supporting doctrines - concerning the purpose of economics, the advantages of a price system, etc - which the theory of marginal-cost pricing rests on. This would require a treatise covering most of the relevant fields treated in books and academic articles on the principles of economics; and this task is not undertaken here. For those interested in such a course, a full bibliography on marginal-cost pricing is providing at the end of this volume. This volume assumes a large area of common ground (such as recognition of the need to satisfy human wants and of the usefulness of money, markets, and prices in measuring and comparing utility to different persons), and attempts merely to show that marginal cost pricing is superior to the major alternatives based upon the same assumptions, especially average-cost pricing and arbitrary 'conservation' pricing, as a theory of efficient price and output.

Besides the introduction there are five sections and three Appendices. They are listed below:

Efficiency and Marginal-Cost Pricing

This section defines economic efficiency and provides the justification for marginal-cost-pricing output control. Examples are given illustrating how marginal-cost pricing leads to an efficient allocation of resources.

The Application of Marginal-Cost Pricing to Electricity

The objective of efficiency in pricing and marginal costs is discussed in the context of electric utilities. Customers and cost causation, long versus short-run marginal costs in pricing, and some qualifications are the other main topics of this section.

Time-of-Use Pricing

This section looks at time-based cost differentials in electric utilities and the justification for time-of-use prices based on marginal costs. The peak problem is set in some historical perspective. Such issues as the measurement of long-run marginal capacity costs when there are many different types of capacity, identification of the peak period, and the allocation of marginal capacity costs to time periods are discussed in detail.

Marginal-Cost Pricing and the Revenue Requirement

This chapter shows how to meet the objective of efficiency in the face of a revenue requirement constraint. The two basic methodologies outlined are multi-part pricing and the inverse-elasticity rule.

Arguments Against Marginal-Cost Pricing

The most influential arguments against marginal-cost pricing are summarized and reviewed.

Appendix I: Efficiency and Marginal-Cost Pricing: A Mathematical Demonstration

This appendix formally defines the objective function of a public utility as maximizing net social benefits after taking account of social costs, and proceeds to show that marginal-cost pricing achieves this goal.

Appendix II: Efficiency and Fairness

This appendix shows that one can meet the objective of efficiency and still meet the constraint of fairness.

Appendix III: A Select Bibliography of Marginal-Cost Pricing

This appendix lists the chief support documents.

II. EFFICIENCY AND MARGINAL-COST PRICING

The object of selling at marginal cost is to provide a valid guide to help users in their choice, in such a way that the least-cost solution for them is also the least-cost solution for society. - Boiteux

If society's scarce resources are to be allocated efficiently among the many alternative uses, the price of a commodity must be based on its marginal costs of production. This is the most important single policy prescription of micro-economics. An increase in the consumption of a commodity produces a benefit to the consumer. Expanding output however, entails withdrawing resources from producing some other item. This additional output of one commodity therefore entails a cost to the would-be consumer of forgone alternative goods and services. The general role of prices is to balance benefits and costs at the margin: that is, to assert proper checks and balances on both production and consumption. Prices have two functions, then:

1. To discourage excessive consumption of a commodity, and
2. To induce the desired supply of that commodity.

A. ECONOMIC EFFICIENCY DEFINED

To develop the justification for marginal-cost pricing, one must first define efficient allocation. Economic theory defines an improvement in the allocation of society's resources as an economic reorganization of those resources, involving a change in the collection of goods produced and in their distribution (through the price mechanism), which could make some people better off (in their own estimation) without leaving anyone worse off. An efficient[1] allocation of resources is therefore one from which no economic reorganization can result in an improvement in resource allocation. Alternatively, an efficient allocation of resources may be interpreted as one that does not contain any 'slack': that is, one where there is no way to reorganize production and distribution that would make anyone better off without making anyone worse off.

Such an efficient state of the economy has associated with it a well-known property: that the collection of finished goods valued at prevailing prices has a higher value than that of any alternative collection of goods that could be produced with the existing resources of society. This property follows logically from the definition of efficiency in resource allocation. Assume, for argument's sake, that it did not, and the point is easily established. If one could reshuffle the existing resources and so produce a yet greater value at the original prices, then it would be possible to give everyone the same value of goods as he enjoyed before and still have some goods left over. The value of these goods left over could then make one or more persons better off. But this implies that an improved allocation is still possible. Therefore, the so-called efficient position was not, after all, efficient as defined. It cannot, then, be otherwise than as was stated: an efficient resource allocation has the property that, valued at prevailing prices, no other collection of goods producible with the same total resources of society could be worth more than the efficient allocation.

Using this framework, the optimal or efficient output level of any commodity may be defined as follows: The amount of society's scarce resources devoted to producing any commodity (say electricity) should be such that the marginal social benefit derived from the last unit of consumption equals the marginal social costs of producing that unit. This proposition is demonstrated with a simple mathematical model in Appendix I of this Volume.

Basically this is so because if the rule is violated, some change in output can benefit society. For example, if the marginal cost of producing one more unit of electricity exceeds its marginal benefit, then society will gain by producing less electricity.

B. MARGINAL-COST PRICING AND THE EFFICIENT ALLOCATION OF RESOURCES

The purpose of marginal-cost pricing is to control the production of a commodity and so allocate resources efficiently.

Marginal cost can be defined very generally as the cost of resources which a society must use to produce one additional unit of some commodity, or the value of resources that it would save by producing one unit less. As long as the goal is economic efficiency, the notion that price should equal marginal cost is a general economic principle having nothing in particular to do with electricity. The principle derives from the basic operation of an economy where decisions about production and consumption are decentralized. Consumers decide how they will divide their incomes among different commodities by looking at the relative prices. Prices act as signals to consumers, indicating the cost to them of using more of various commodities. Insofar as commodity prices equal the marginal social costs of production, these price signals indicate simultaneously the cost of commodities to individual consumers and the cost of producing such commodities to society as a whole. With prices set equal to marginal cost, consumers' decisions about the trade-offs associated with the consumption of different commodities are guided by signals which reflect the actual trade-offs in resource costs associated with producing those commodities.

For example, if the price of some commodity such as electricity is set below its marginal cost, consumers will think an additional unit costs less than it really costs society to produce it. The consumer will be led to increase his consumption until the marginal value of an additional unit of the commodity equals its price. But since the price has been set below the marginal cost, the value of the last unit of consumption to the consumer is less than what it costs society to produce it. More resources are being devoted to the production of this commodity than is socially efficient.

Society's resources will be better allocated if price is set to equal marginal cost. Then the consumer will still increase his consumption where the marginal benefit of an additional unit of the commodity equals its price. And the price will reflect the marginal cost of using society's resources to produce another unit of the commodity - an ideal allocation of resources devoted to producing that commodity. There is neither an underallocation nor an overallocation. The level of output is optimal, in that the added value of the last unit of production equals the added cost of producing it.

Why does economic efficiency require prices equal to marginal cost, instead of (for example) average total costs? The reason is that demand for all goods and services responds to some degree to price.[2] If consumers are to decide honestly whether to take a little more or a little less of any particular item, then the price they have to pay for it (and the prices of all other goods and services with which they compare it) must reflect the cost of

[1] The terms 'optimal', 'Pareto optimal' and 'ideal' may be used interchangeably with 'efficient' here. Pareto was the economist who first derived the efficiency conditions. (See V. Pareto, *Manuel d'Economie Politique*, 1909).

[2] See Volume IV, *The Demand Elasticity Study*, which evaluates consumers' responsiveness to the price of electricity.

supplying a little more or a little less: in short, marginal cost. If buyers are charged more than marginal cost for a particular commodity (for example, because the seller has monopoly power), then they will buy less than the socially efficient quantity. Consumers who would willingly have had society devote the incremental resources required will refrain from making those further purchases, because the price to them exaggerates the sacrifices.

Conversely, if price is below marginal cost, then production of the product in question will be higher (and that of all other products taken together, lower) than it ought to be. Society is sacrificing other goods and services to produce more of the particular item than customers would willingly have authorized had the price to them fully reflected the marginal cost of production.

C. MARGINAL-COST PRICING: EXAMPLES

Mining and farming provide useful illustrations. Consider coal, where added production has to come from less favourable sites and sources, where the cost of producing one more unit is higher than under the original more favourable or "intra-marginal" conditions. Here there is an obvious reason for making this higher or marginal cost the basis of calculation, since this will represent the true social cost of expanding production slightly. (Or, alternatively, the cost that is saved by contracting it slightly.) If any further coal produced must come from working less accessible or less productive seams, then cost under these more difficult conditions is alone relevant to the choice between coal and rival fuels.

The same is true of farming, if growing more grain means tilling poor soil, or soil not suited to grain. This idea is one of the oldest in economics,[3]
and forms the basis for the classical theory of rent.

If in these circumstances coal or grain were priced by averaging the cost under more favourable and less favourable conditions, this would in effect use the rent from the former to subsidize an uneconomic expansion of the latter. Clearly, then, the average-cost price encourages producing coal or grain at a high cost at the margin, rather than using cheaper substitutes.

This point may be illustrated by a fable attributed to Gabriel Dessus:

> *Let us imagine that a village set at the foot of a wooded mountain owns (among other things) a coalmine. Let us suppose the villagers do not care whether they are woodcutters or miners. For heating they will first of all resort to cutting wood at the foot of the mountain; then, as their need for heating grows, they will begin to exploit the woods higher up, until the day comes when they perceive it as less of a burden to go and dig coal from the mine than to go and exploit the high forest. Equilibrium between the prices of the two fuels will be reached at the moment when the marginal cost of woodcutting equals the cost of coalmining.*
>
> *All this is completely classical. But if the community decides to put the sale of fuel under public control, what pricing-policy should it follow? If it follows a policy of sale 'at cost' (that is to say, at average cost), it will have to lower the price of wood compared to its price in the previous competitive situation, since average cost is visibly less than marginal cost; while the price of coal will remain unchanged. In these circumstances, the inhabitants will buy a little more wood, and the community will be forced to exploit the forest a little higher in the mountain still; which amounts to saying that by its pricing policy, the village will have obliged its citizens (in their role as woodcutters) to accept a useless increase of effort. The sensible policy would be to sell the wood at its marginal cost (and coal also, but we have assumed that the cost of extracting the latter is constant). The municipality will have to resign itself to making some profits, for which it will surely be able to find a use.*

Ultimately, then, there is a common element to all decisions which can be expressed in the apparently trivial question, Is it worthwhile? A firm considering whether to improve the quality of a product, or a consumer debating whether to buy a bottle of wine must both ask the same question: will the action in question add benefit enough to make it worth the cost? This is the heart of marginal decision-making: that an action is worth taking if, and only if, the actor can expect to be better off from it than he was before.

A special case of the general theory that price should equal marginal cost is perfect competition for businessmen: it pays to produce and sell so long as incremental revenues cover incremental costs. In a competitive market, incremental revenues are simply defined as the market price times the added quantity sold. Hence the elementary proposition that under perfect competition businessmen will increase production until their marginal costs equal their price. Competitive behaviour, then, assures the equivalence of price and marginal cost that is required if free consumer choices are to yield an efficient allocation of resources.

[3]See David Ricardo's classic statement on rent in *Principles of Political Economy and Taxation*, Chapter II.

III. APPLYING MARGINAL-COST PRICING TO ELECTRICITY

A. THE PRICING-OBJECTIVE AND MARGINAL-COST PRICING

It is important in this discussion to keep in mind the recommended objective of pricing for electricity:

The pricing-structure should help to efficiently allocate resources used in producing electricity.

Given the objective of efficiency, then, it is almost a matter of definition that the price of electricity ought to reflect the marginal cost of producing it. A rate structure based on marginal costs may be viewed as part of an information feedback loop which leads to the lowest-cost system of producing electricity through time. The customers receive the marginal-cost information through the rate structure. They then make their own cost-minimizing (satisfaction-maximizing) calculations about consumption and willingness to pay. This information, in terms of time of use and rate of use, feeds back to the system planner. The feedback gives him the relevant demand information about load growth and optimal generation mix. It is the relevant demand information because the customers' decisions are based on prices that reflect the marginal cost of production.

The objective of efficiency and marginal-cost pricing would be consistent with the corporate objective, which is

To supply the demands of the people of Ontario for electric energy at the lowest feasible cost consistent with safety to its employees and public and a high quality of service to its customers, and subject to the social, economic, and environmental concerns of the people of Ontario.

The corporate objective implies, then, that Ontario Hydro should operate on its least-cost curve through time, neither underbuilding nor overbuilding. To do this requires the optimal amount of generation and optimal generation mix based on customers' demands. What is meant by optimal? It means that if the price the customer faces reflects the marginal costs associated with producing electricity, then the customers will reveal through their decisions about use - their dollar votes in demanding electricity - just how much plant to build and how much energy to produce.

Another way to put this is that the Corporation will maximize social benefits net of social costs. This is achieved by giving the customer the right price signal (that is, the marginal-cost information) through prices. If the marginal use of electricity is priced at marginal cost, then resources devoted to producing electricity will be allocated efficiently. Marginal-cost pricing yields an optimal level of production because the marginal social benefits the customers derive from electricity will equal its marginal social costs of production. It is impossible in this situation to increase net social benefit by increasing or decreasing production through price adjustments.

In meeting the objective of efficiency, there are two constraints imposed. These are:

1. The revenues generated through marginal-cost pricing should meet the revenue requirement in any given year, neither more nor less;
2. In meeting the objective of efficiency the price structure should be as fair as is practicable.

Section V pursues the first constraint in dealing with marginal-cost pricing and the revenue requirement. In the second constraint 'fair' implies equal treatment of equals, based on criteria which reflect general agreement in the community. These agreed criteria of equitability may be listed as follows:

1. The pricing-structure must maintain the integrity of the cost-pooling concept.
2. There should be no seniority rights in pricing-structure. All consumption is always new, for the customer may decide to discontinue it at any moment.
3. The price structure should be impartial. There should be no discrimination; all end users should be priced at the same rate.
4. The pricing-structure and changes in the price level should be defined clearly, so that the customer knows what price he will pay if he undertakes a specific course of action. This is the criterion of certainty in prices.
5. Changes in corporate policy (for example, level of system reliability) should not lead to abrupt changes in quality of service received or price charged. This is the criterion of continuity in prices.
6. When costs are increasing, the economic quasi-rent known as 'the historical benefits of investment' should be apportioned among customers impartially. That is, the electric utility's pricing-structure should be distributionally neutral. Similarly, when costs are decreasing, the economic quasi-rent known as 'the historical burden of investment' should be apportioned impartially. This is known as 'the criterion of distributional neutrality', and simply means the utility's rate structure should not be used for 'social engineering' or redistributing wealth.

The subject of fairness has had a long history in rate-making. A more formal discussion on the appropriate approach to fair pricing is found in Appendix II: Fairness and Efficiency. Further discussion of the traditional objective of fairness and pricing is found in Volume VI.

Essentially, then, the purpose of marginal-cost pricing is to ensure that the cost a customer incurs or saves in making a decision about use reflects the cost incurred or saved by the utility, and ultimately by society, in providing the commodity. While it is important for prices to be fair, it is also important for the concept of fairness to be defined with the same rigorous logic as the objective of efficiency.

One must take care, though, not to overstress fairness as a pricing-goal. If a rate structure required to obtain full efficiency is held to impose an unfair burden on any particular group, then methods can often be found to compensate them in other ways, without the impairment of efficiency inherent in attempting to distort rates to conform to the ratemakers' notion of fairness. If, on the other hand, efficiency is lost through the rate structure, then there will in general be no way to make up for the loss. In any event, it is important to keep in mind, in judging the fairness of a proposed rate change, that fairness is a matter ultimately of the total effect on individuals.

Often fairness and efficiency will point in the same direction. Indeed, there is a *prima facie* presumption that a rate structure that comes closer to reflecting the relevant cost elements will be fundamentally fairer than one that distorts them.

B. CUSTOMERS AND COST CAUSATION

Marginal cost is the cost a customer imposes on the corporation and society to provide him with each unit of the product he consumes. It is important to note that all consumption is always new, for at any moment the customer can decide to discontinue it. That is, at each moment, each kilowatt-hour is a new kilo-

watt-hour. It matters not that a certain customer may have been located beside Sir Adam Beck Niagara GS No 1 since 1925. No seniority rights accrue to any customer or customer class. As far as responsibility for cost is concerned, every demand is new. It is important to recall too that it is always the cost of new plant that must be considered when a customer decides to reduce, maintain, or increase his use.

Alfred E. Kahn elaborates on this approach to marginal-cost pricing.

> *Suppose, for example, the utility has two groups of customers: one, A, whose demand is stable; another, B, whose demand is increasing. And suppose expansion of the latter demand finally requires expansion of capacity. Does that mean, following our rules of peak responsibility pricing, that B are the marginal buyers on whom capacity costs alone should be imposed? Obviously not. True, it is the increase in B's purchases that precipitates the additional investment; but the additional costs could just as well have been saved if A reduced their purchases and if B refrained from increasing theirs. So A's continuing to take service is just as responsible, in proportion to the amount they take, for the need to expand investment as B's increasing needs; and A should therefore be forced just as much as B to weigh marginal benefits of the capacity to them against the marginal costs they impose on society for continuing to make demands. This reasoning applies even when incremental investment costs per unit of capacity are rising, and where, again, it might appear on first consideration that since it is the expansion of the B demands that is responsible for the suppliers' incurring the higher costs, it is that group that ought to bear the additional burden. Even though B's demand is marginal in a temporal sense, both groups are marginal in the economic sense.*[4]

C. LONG-RUN AND SHORT-RUN MARGINAL COSTS

Although the case for marginal-cost pricing is clear, the question remains whether long-run or short-run marginal cost is relevant for the pricing-rule. Obviously, for a system that has been perfectly adjusted to demand given prevailing technology there is no issue, since long-run and short-run marginal cost are equal.

This point has been demonstrated by Ralph Turvey, among others. The traditional theory of cost and supply curves starts by taking as given the technical conditions of production and the supply conditions of inputs to production. With the production function and input supply curves thus established, it then postulates cost minimization and derives the minimum-cost input combination for each set of alternative production levels. For a single-product firm, the result is expressed as a long-run average-total-cost curve relating annual cost to annual production.

Short-run average-total-cost curves are derived by supposing the input of one factor to be fixed, each short-run-cost curve corresponding to one amount of the fixed factor. Any such curve cannot lie below the long-run cost curve at any point; for then the long-run curve would not be the minimum-cost solution. But such a curve can very well rise above the long-run curve for part of its length, since there can be some outputs for which the minimum-cost input combination includes more or less of the fixed factor than the fixed amount. At outputs where this fixed amount coincides with the optimal amount, the two curves necessarily coincide: short-run and long-run average costs are equal.

At the level of output where the two average-total-cost curves coincide, the two corresponding marginal curves will also coincide. Hence the theorem that short-run marginal cost equals long-run marginal costs at those outputs where the actual amount of the fixed factor coincides with the optimal amount. If the fixed factor is called 'capacity', then this theorem can be restated by saying that marginal short and long-run costs coincide when capacity is optimal. This is an important theorem, because it shows that the argument about whether public enterprises should set prices equal to long-run or short-run marginal costs is only meaningful when capacity is not optimal.

It should be noted that the definition of short-run marginal cost includes "all sacrifices, present or future, external as well as internal to the company, for which production is at the margin causally responsible".[5] Consider the marginal costs of generating electricity, for example. In calculating short-run marginal cost for any hour, one would have to consider the following costs:

1. running-costs,
2. external costs (for example, pollution such as atmospheric emissions from coal-fired power plants); and
3. "the cost of not having enough".

Both fuel costs and external costs associated with generating electricity are familiar enough. The cost of not having enough needs more explanation. It corresponds to the costs incurred by customers who would not be served if demand exceeded capacity.

These costs are called 'curtailment costs'. In principle, curtailment costs could be calculated directly. The French nationalized electricity industry does just that: it looks at its plan for load-shedding and calculates the loss in value added for industries which it would shed if threatened with a power failure; it then plans to add capacity until the cost of the last unit of capacity added equals the expected cost of a failure. In other words, for an optimally designed system, capacity is added up to the point where the marginal cost of capacity equals the expected marginal cost of curtailment.

The long-run marginal cost is measured *after changes in capital equipment* have been made to adjust to increased demand in the long run, and takes into account the capital costs of increased production.

With proper planning, the long-run marginal costs will equal the short-run marginal costs. This may seem odd, but on reflection it is really not odd at all: if the cheapest way to meet increased demand is to run existing machines, then they will be run to the point at which the extra cost of running them, plus possible shortage costs, equals the cost of meeting the extra load by changing the capacity. If the extra running-costs, plus shortage costs, rise above the point at which it is worth substituting capacity for running-costs, then if the planner has anticipated this, he will meet his goal of minimum cost by putting in the extra capacity.

In North America, the reserve margin for generation and transmission is decided by reference to a set of reliability criteria, and it is assumed that these criteria adequately reflect the costs of not having enough, or the curtailment costs. It is assumed, then, that the system has been planned to the level of capacity where

[4] A.E. Kahn, *The Economics of Regulation* (1970), vol. I, p. 140.
[5] Kahn, *Economics*, vol. I, p. 75. See Volume VI, Section IV, for a more detailed analysis of external costs.

the marginal cost of capacity equals the expected marginal cost of curtailment. The curtailment cost can be determined (as it were) in reverse by looking at the capital cost of the last unit of capacity. One must take some care in defining 'the last unit'. Generally it is the lowest-cost generating-unit which the utility would use to meet the peak. In other words, the marginal unit-curtailment cost equals the annual marginal unit capital cost per kilowatt on a peaking-unit. This is the optimal case, then, where short-run marginal cost equals long-run marginal cost.

For example, if growth has the same load-duration curve as the current system, then the capital equipment needed for the optimal system discussed above is reproduced in miniature for the increment. The sum of the short-run marginal costs equals the sum of the marginal costs when the system has fully adjusted. It also holds for each set of hours.

Suppose growth occurs only at the peak. Then only peaking-plants will be added, and the price of running-cost and capital cost of a peaker at the peak will exactly recover the total costs of adding and running the peaking-plant.

If the growth is expected to occur in the off-peak hours, the system plan has to be re-optimized. The new optimal plan will include more baseload and less intermediate capacity. If this has been accurately forecast, the extra capital cost of the baseload plant will be exactly offset by the reduced running-costs in the intermediate period. This is shown in Corollary B of Appendix I in Volume VII.

In the real world, plant may not be perfectly adjusted. Since plants take quite some time to build, and there is uncertainty about what demand and production costs will be when they finally come on line, there may be, at any time, a mix of capacity that is less than optimal. Theoretically, in such situations price should always equal short-run marginal cost, in order to obtain the most efficient use of the existing plant. When there was substantially more capacity than was optimal, this would mean prices below long-run marginal cost. When there was less capacity than was optimal, this would mean prices above long-run marginal cost, high enough to constrain actual demand to available capacity.

A more likely situation would involve a sub-optimal *mix* of plants. In many parts of North America, substantial oil-burning baseload and intermediate load capacity was built on the assumption of cheap oil. Now that oil is much more costly, the mix of plants and how they are used over time will gradually change as new equipment is added. So while marginal energy costs during off-peak periods are quite high today, they may be lower in the future as new capacity is added. Prices based on short-run marginal cost would fall over time, as plant mix changed. In this example, then, the long-run marginal cost of off-peak power would be substantially below the short-run cost.

Thus, one result of prices based on short-run marginal cost would be considerable fluctuation in rates over time, reflecting changing levels of reserve capacity and changing mixes of plants and fuels. In principle, this does not present any real problems, but in reality, it could cause both serious administrative problems and serious economic distortions, especially with electricity. For one thing, it might be very difficult and costly to have to keep revising rate structures to reflect short-run marginal costs.

Perhaps more important are distortions which may result. Customers' decisions are largely appliance decisions. Consumers must make long-term investment decisions when choosing appliances. Such decisions must be based on price expectations about energy that are highly uncertain. If consumers had the best information available about what prices were likely to be in the future, and if there existed perfect insurance markets to help them diversify the risk associated with durable goods purchased in an environment of uncertain energy prices, distortions from short-run marginal-cost pricing would not exist. However, neither situation is likely to hold. The companies have much better information about what their long-run costs will be than consumers. Moreover, there are few ways consumers can diversify the risk associated with uncertain prices for energy.

Prices based on long-run marginal costs, then, would promote rate stability and provide consumers with good long-run signals for making intelligent decisions about electric appliances. Where short-run marginal cost is not likely to be very far from long-run marginal cost over the planning period of from five to ten years, prices should be based on long-run marginal costs. The costs of short-run distortions are likely to be small compared to the saving in administrative costs and the value of the information such prices give consumers. Prices should therefore be based on long-run marginal costs until cost-benefit analysis suggests there would be a net gain in efficiency from using short-run marginal costs. And again, once the optimal generation mix has been achieved, the issue becomes a mute one, because long and short-run marginal costs will be the same.

Further discussion on the marginal cost of producing electricity will be found in Volume VII: *Costing-Methodology for Determining Marginal Costs*. The operational definition of marginal cost is given there, along with the justification for the costing-methodology, the step-by-step costing-procedures, and the bottom-line cost results.

D. SOME QUALIFICATIONS

Proposing marginal-cost pricing to promote efficient allocation of resources devoted to producing electricity implies several assumptions.[6] They are:

1. Since the distribution of real wealth is not properly the business of a public enterprise, it should act as though distribution were always ideal.
2. The customer is always right.
3. There are, unless specified, no externalities either of production or consumption.
4. What is not known should be ignored.

These assumptions taken together imply first that the social benefit a public enterprise generates is measured by its customers' willingness to pay for its outputs and second that its money costs differ from its social costs only for specified externalities and known divergences between the price of inputs and the value of their marginal products in other uses. Each assumption merits further comment.

1. The Distribution of Real Income Is Not the Concern of a Public Enterprise

Ontario Hydro's mandate is to meet the electrical demands of the province at the lowest feasible cost. Achieving that objective requires an efficient pricing-system. Such a pricing-system necessarily conveys to the customer the costs the utility is about to incur on his behalf to produce electricity. Social engineering, whether through rates or through deliberately letting some cus-

[6]See Ralph Turvey, *Economic Analysis and Public Enterprises* (1971).

tomers cross-subsidize others, is not within the Corporation's mandate. Furthermore, the government should not use Ontario Hydro as a policy instrument in order to redistribute wealth.

The pricing-objective of redistributing wealth is closely analysed in Section III of Volume VI, *Alternative Pricing-Philosophies and Rate Structures*. Such pricing-schemes as lifeline rates, and energy stamps received detailed examination. It was found that there were major drawbacks to any program for redistributing wealth that Ontario Hydro might undertake. In summary, these were

1. Inefficient allocation of resources used to produce goods and services within society.
2. Inefficient allocation of resources devoted to producing electricity.
3. Inefficiency in identifying and helping low-income individuals and families cheaply.

2. The Customer Is Always Right

This assertion is not as strong as it might first appear. It means only that an individual has a preference function, in that he chooses among the alternatives open to him to obtain the greatest possible satisfaction from consuming commodities (in the broadest sense). This implies that he knows the alternatives facing him and can evaluate them.

While there may be people who cannot judge their own interests competently, that is an argument for education rather than paternalism.

For example, the energy efficiency should be clearly stated for all appliances sold. The cost of consuming one more kilowatt-hour or one less should be clearly stated on the customer's bill. Municipal utilities and other suppliers of energy (such as gas distributors) could provide consumers with total-cost figures for alternative heating-methods, thereby letting residential users compare the alternatives in terms of their present values.

3. There Are No Externalities, Unless Specified, Either of Production or of Consumption

Economic efficiency requires a 'shadow price' on production activities that give rise to an externality. Such a shadow price should equal the marginal net damage caused from producing electricity. Price would then equal marginal social cost. The customer would then face a price which represented the real resource cost consequences of consuming one more kilowatt-hour or one less. Shadow pricing would be fully consistent with the objective of efficiency.

However, there are two main drawbacks to shadow pricing at present:[7]

a. Information Requirements

As was said, the proper shadow price equals the marginal cost of damage from producing electricity. At present, it is practically impossible to obtain a reasonable estimate of the money value of this marginal damage. That is, many of the most important consequences involve such things as damage to persons' health or to the beauty of the countryside, which are difficult to estimate in monetary terms.

b. Inflexibility

One drawback to a shadow-pricing scheme is implicit throughout all discussions: its lack of flexibility. Shadow prices would be hard to change on short notice, or to implement on a regional basis. They do not allow for differences in the effects of equal quantities of emission upon the effective level of pollution. Such inflexibility makes shadow pricing a questionable tool for dealing with the external costs associated with producing electricity.

In the light of the foregoing analysis, shadow pricing, or an externality tax, should be rejected as a policy alternative. The Corporation should base its prices on marginal private costs, while meeting the environmental standards and criteria set by government.

In meeting prescribed standards, of course, external costs are internalized. After internalization is achieved, the Corporation would consider all costs it was about to incur in setting prices based on marginal costs.

There are effective alternatives to shadow pricing: prescribed air quality standards, emission standards, pollution charges, and direct controls to name the main ones. Some degree of arbitrariness in the design of such standards is inevitable, it is true; moreover, such a route gives up any attempt to reach a true social optimum in a theoretical sense. To follow these alternatives is essentially a 'satisficing' approach[8] to the problem; yet it does offer some significant optimality. Besides offering administrative savings and being comparatively easy to carry out the above measures, if properly designed and implemented, could lead to attaining the chosen standards at approximately the minimum cost of society.

4. What is Not Known Should be Ignored

In essence, this assumption means that it is not within a public utility's mandate to take account of resource misallocations elsewhere in the economy when making decisions about pricing and production. This means that second-best pricing considerations should be co-ordinated by government.

Obviously, the decision to price at marginal cost in one market strictly depends on the relationship of price to marginal cost in other markets (primarily markets for products that are good substitutes for the one under consideration). If prices in other markets deviate from marginal cost and cannot be changed, deviations of price from marginal cost in all other markets may be called for to preserve efficiency. In taking such considerations into account in pricing, the starting-point must still be marginal cost.

[7]There are seven important drawbacks to shadow pricing discussed in Section IV of Volume VI. The two dealt with here are the most significant.

[8]That is, there is no attempt to seek any sort of optimum. Instead one seeks to merely find policies that will meet some present standards, and so produce results considered acceptable or satisfactory.

IV. TIME-OF-USE PRICING

A. ECONOMIC CHARACTERISTICS OF ELECTRICITY

There is one basic feature of its market which gives electricity (together with a few similar goods) a special place from the point of view of pricing. Generally it is not economically feasible to store electricity in significant amounts. Since demand for electricity fluctuates over time, this means that the capacity of each piece of capital equipment (generating-stations, transmission lines, transformers, etc.) is determined by the highest demand which that particular piece of equipment is expected to meet.

Because the demand for electricity is periodic, varying by time of day and season of the year, and its supply basically cannot be stored, the costs of supplying additional consumption also vary. Taking additional consumption when capacity is fully or almost fully utilized normally entails running peaking-plants, which have high energy costs, longer. Additonal consumption in the peak period may also (given plant of fixed capacity) cause a shortage in which some demand cannot be met. In the longer run, where plant can be varied, a sustained increase in load at those peak hours may require the utility to add more capacity.

When demand is low compared to available capacity, one can meet additional demand by running plants with relatively low running-costs longer and still run minimal risk of a shortfall in capacity. In the longer run, a sustained rise in off-peak consumption would lead a utility to change its mix of plants on line, to take advantage of the generation economies associated with a higher load factor and yet not increase the overall capacity of the system. Added consumption off the peak, even if sustained, does not impose added capacity costs on the system. It is therefore generally conceded that marginal costs are usually higher during peak periods than off peak. Pricing-systems based on marginal costs which vary by time of day and season of the year have generally been called systems of 'peak-load pricing'. It must be remembered, however, that peak-load pricing is nothing more than the application of the principles of marginal-cost pricing to a situation in which marginal costs vary by time of day and season of the year.

There is no unique definition of the output of an electric utility. There are, however, four characteristics of the output which affect the utility's costs and demand for the product. For each characteristic of these, the consumer faces a decision about use. The four key decisions are:

1. Decision to Become a Customer

There are identifiable costs associated with the customer's decision to connect with the system and continue as a customer. These may be called customer costs, and lead to a customer charge.

2. Decision about Rate of Use

There are identifiable costs associated with the capacity requirements of the utility (cost per kilowatt) which result from customer demands about rate of use. These may be called demand costs, and lead to a demand charge.

3. Decision about Use

There are identifiable costs which vary according to the energy requirements of the utility (cost per kilowatt-hour) which result from customer decisions about use. These may be called energy costs, and lead to an energy charge.

4. Decision about Time of Use

There are identifiable costs which vary according to the time when energy requirements face the utility. Generally this means there is a cost per peak kilowatt-hour different from the cost per off-peak kilowatt-hour. This cost differential leads to a rate differential between peak and off-peak energy use. In addition, it provides a basis for the assignment of the demand charge.

B. TIME-OF-USE RATES VIEWED HISTORICALLY

The problem of the peak was first approached by the use of a two-part rate, credited to John Hopkinson, an English engineer. Today, it is almost universally used by electric and gas utilities for large-volume sales in bulk and to industrial users. The Hopkinson two-part rate may be described as follows. The first part - the energy or commodity charge - embodies the variable costs, properly charged to all customers. It is levied by unit of consumption (per kilowatt-hour or per millions of cubic feet (MCF) of gas). The second part - the demand or capacity charge - is a charge for the utility's readiness to serve on demand. This readiness to serve is made possible by installing of capacity. The demand charge thus distributes the costs of providing the capacity, the capital costs. With the Hopkinson rate, the demand charge was assessed on the basis of the customer's maximum demand in a given period (e.g., a day, a month, a season).

The Hopkinson two-part rate was hailed as a great discovery at the time. Unfortunately (as Alfred Kahn has noted), it was based on a simple confusion. If the demand charge were correctly to reflect peak responsibility, then it would impose on each customer a share of capacity costs equivalent to his share of total purchases at the time coinciding with the *system's* peak.[9] Instead, the typical two-part rate bases that rate on each customer's *own* peak consumption over some measured time period, regardless of whether *his* peak coincides with that of the system. Hence the designation "non-coincident" demand charge. That is, the peak (for example) half-hour consumption of all customers, regardless of the time of day or year in which each falls, is added up, and each then is charged a share of total system capital costs equivalent to the percentage share that his peak consumption constitutes of that total.

The non-coincident peak method as traditionally applied does not track costs; it merely allocates them. What measures the share of capacity costs for which each customer is causally responsible is his share of the use at the system's peak; for consumption at that time decides how much capacity the utility must have available. The system's load factor might well be improved by inducing individual customers to cut down their consumption to a deep trough at the system peak and enormously increase their own peak use at the system's off-peak time: yet the non-coincident demand system would discourage them from doing so. The point is a simple one. The maximum rate at which the individual user takes his power is irrelevant; what matters is how much he takes at the time of the system peak.

If the two-part rate is to work satisfactorily, the demand charge each individual pays really must bear a fairly close relation to the costs he imposes on the system at its peak. In the early days of

[9] This, of course, refers to the simple case where the system peak is clearly and narrowly defined. It should be noted that the marginal cost of capacity cannot be attributed solely to a single hour of peak demand. One can consider the responsibility for capacity a graduated-peak responsibility, and then use the relative value of the loss-of-load probability in each hour to estimate the graduated responsibility for capacity cost. (See Part F and Volume VII for further discussion on assigning capacity costs.)

the industry, when peak consumption was confined to lighting-demands, the non-coincident demand method was a reasonable approximation to peak responsibility. In these days, this is no longer so.

C. THE THEORY OF PEAK-LOAD PRICING[10]

The demand for an electric utility's output varies periodically over the day and over the year. Peak-load pricing deals with the problem of meeting these variations in load with some optimum-sized plant capacity and capacity mix with the accompanying investments and costs, all within the framework of a pricing-structure.

The maximum capacity of the plant and the pricing of the utility's production depend on the set of assumptions used to define the peak-load problem. Consider a highly simplified example in which there are no diverse technologies, but only a single plant.[11]
Assume further, that a single service is to be produced during two periods, for which only two costs are incurred. Let b be the marginal running-cost per unit per period, and let B be the marginal cost of providing a unit of capacity. Assume further that both b and B are constant. It has been shown that in the long run the marginal cost of a unit of output will be b (which is the marginal running-cost) if capacity is not used to the full and b + B (which is the marginal running-cost together with the marginal cost of a unit of capacity) if all the existing capacity is used up. If the demand curves are given for each period, it is possible to determine the efficient level of output in each period and the prices associated with these outputs. The capacity required will be the peak demand on the system. Given the pricing objective of efficiency, the following peak and off-peak prices should apply to each customer.

Peak Charge = (b X peak kWh) + (B X peak kW)

Off-Peak Charge = b X off-peak kWh

Published studies have developed more complex theoretical models, which simultaneously treat periodic demand, diverse technology, uncertainty of demand, and curtailment costs all within a consistent framework for setting efficient prices.

From this simple model, two general principles of efficient peak-versus-off-peak pricing may be enunciated.

1. Set the price by the hour or hours of the day or season of the year, in accordance with the pattern of marginal costs and demand. It should be noted that this principle implies that in any particular hour every customer should be treated the same; this excludes, then, charging users differentially depending (for example) on their volume of use or type of use. Thus it would rule out such things as volume discounts and all-electric rates.
2. No responsibility for capacity costs is imputed to those customers whose demand does not press upon capacity (where the peak is simply and clearly defined). This principle does not spell out in detail how the responsibility for capacity should be allocated among the various peak users. The recommended method in the rate design proposals is to employ the customer's maximum demand in the peak period. The method of assignment is important because instantaneous demand peaks do not necessarily stay fixed. This subject will be returned to shortly.

Some argue against the very principle of peak-versus-off-peak pricing. The three most prominent arguments are:

1. It is unfair and discriminatory to charge peak use alone with capacity costs, since the capacity obviously serves all users at all times.
2. The utility has a special responsibility to protect the householder, and should try to keep his rates down.
3. The utility should promote the maximum extension of its service, subject only to the condition that the aggregate revenues should cover aggregate costs - goals that may well conflict with peak-responsibility pricing.

Arguments such as these are for the most part not susceptible to scientific refutation, since basically they involve non-scientific value judgements. As A.E. Kahn[12] noted, economists and proponents of efficiency can only cite the following counter-considerations:

1. In economic terms, peak responsibility pricing does not discriminate between peak and off-peak users. Discrimination occurs when there are differences in price that do not correspond to differences in costs. It costs more to supply users at the peak than off peak, and the proposal is simply to reflect that cost difference in the respective prices. Every peak user actually imposes on society, in the long run, the incremental cost of the capacity on which he draws. There is no such causal connection between off-peak use and capacity costs: the capacity would be there whether the off-peak user made demands on it or not. It would be truly discriminatory to levy any of these costs on the off-peak user.
2. In this sense, it is unfair to make off-peak users pay some share of capacity costs, for which they are not themselves causally responsible.
3. Moreover, such a policy would be economically inefficient. Insofar as off-peak demand has any elasticity at all, a charge to these users that incorporates any capacity costs will cause them to give up satisfaction the true social costs of which they would be perfectly willing to pay; and some productive capacity is left wastefully idle. Conversely, and subject to the same condition, if peak users do not have to pay the full marginal costs of being supplied, they will induce society to provide them with capacity using resources that would have given greater satisfaction if directed to other ends.
4. Under these circumstances, off-peak users would be subsidizing peak users. Even if society were to make a conscious decision to transfer income from off-peak users to peak users, such a policy would not be an intelligent one, if it did not take into account that departures from efficient pricing would be an economically inefficient way to make such a transfer. The 'proof' of this depends only on peak demands having some elasticity. If the transfer were made as a money grant to those users, instead of in the form of prices below cost, they would not use all that money to buy the utility's commodity, but would spend some of it for various other goods and services.

---

[10]For a comprehensive theoretical treatment of peak-load pricing see the "Symposium on Peak Load Pricing", *Bell Journal of Economics and Management Science*, Spring 1976.
[11]For a more detailed model see Appendix II of Volume VII, "A Simplified Model of Time-of-Day/Seasonal Pricing" by Sally Hunt Streiter and Leo T. Maloney, Jr.
[12]*Economics*, vol. I, pp. 100-103.

D. THE MEASUREMENT OF LONG-RUN MARGINAL CAPACITY COSTS WHEN THERE ARE MANY DIFFERENT KINDS OF CAPACITY[13]

One issue that has caused some problems is the notion of a long-run cost of capacity when there are many different kinds of capacity with very different construction costs. As will be shown, there really is no problem.

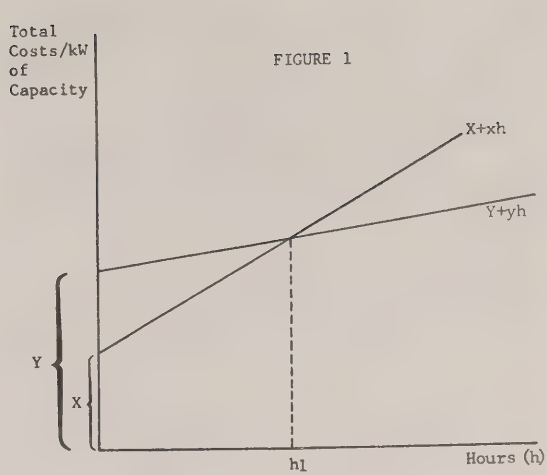
Generally, the system planner faces a whole menu of alternative mixes of generating-plant, each with different construction and running-costs. Some plant have high running-costs but low capital costs, while other kinds have relatively low running-costs but high capital costs. The average cost of the first kind of plant is lower than that of the second, if they are to be run for only a small part of the year. The costs per unit of the second type of plant fall below those of the first if they are run for a substantial part of the year. Since electrical load is periodic, requiring some capacity for short durations and some for long, normally there will be some plant of each kind in the generation mix.

The measurement of the marginal capacity cost turns on this decision. Consider first a system that had only plants with low capacity costs but high energy costs (peaking-plants). Such a system could reduce the total costs of producing power by substituting some plants with high-capacity cost and low energy cost (base loads) for the peaking-plants, since it could use these to meet the loads of longer duration more cheaply. Total system costs would be reduced as long as the difference between the cost of constructing another baseload plant and the fuel savings from replacing a peaking-plant were smaller than the cost of constructing the peaking-plant. As more and more peaking-plant was replaced with baseload plant, the marginal savings in fuel costs would decline. The optimal mix of the two would exist when the cost of an additional unit of baseload capacity, less the marginal fuel savings from this addition, equalled the cost of an added unit of peaking-capacity. Thus for a minimum-cost system, the long-run cost of capacity for any type of plant would be simply the marginal capacity cost less the marginal fuel savings from an increment of such capacity, *given a fixed load-duration curve.* The least-cost combination of plant would be achieved when the marginal construction costs less the marginal fuel savings were equal for all kinds of plant. Since peaking-plants have the highest energy costs, there are no fuel savings associated with more or less of this type of capacity; therefore for a least-cost system, the marginal cost of capacity is simply the cost of constructing peaking-plant. A very simple rule of thumb is that the marginal cost of additional kilowatts of capacity is the marginal cost of peaking-capacity.

This rule of thumb can be shown algebraically. Assume two types of plant, a peaking-plant (with low capital costs but high running-costs), and a baseload plant (with the reverse), and use the following symbols:

	Annual Capital Cost $/kW	Running-Cost $/kW
Peaking-Plant	X	x
Baseload	Y	y

The above information can be displayed graphically as in Figure 1.



In an optimal system, peaking-plants are used when $X + xh = Y + yh$. Since the system planner knows X, Y, x, and y, then $h_1 = (Y-X) \div (x-y)$.

Now the fuel savings from the baseload offset part of the capital cost of the plant, so that the net cost per kilowatt equals the cost of the peaking-plant. That is, the capital cost of the peaking-plant equals the capital cost of the baseload plant *less* fuel savings from running the baseload rather than the peaking plant:

$(Y-X) \div (x-y) = h_1$ and $X = Y - h_1(x-y)$.

In reality, generation planning is a far more complex undertaking than that described above. Unforeseen changes in costs, and institutional barriers, may make a truly minimum-cost system unattainable at any given time. However, assuming that firms have scheduled their construction plans to minimize the production costs given some expected level of demand, then the construction plan can be used not only to meet loads but also to get an even more refined estimate of long-run marginal cost.

It is important to note that reserve margin is included when determining marginal capacity cost. When an electric utility makes plans to construct plant, it aims to build sufficient margin to allow for contingencies. These include the need for maintenance, the possibility of forced outage on machines, and the likelihood that demand may be somewhat higher than predicted because of random factors such as a particularly hot summer or severe winter.

The (correct) reserve margin should not be thought of as excess capacity. It is there to be used. The higher the reserve margin, the greater the *reliability* with which a given kilowatt-hour is provided. The decision about the correct reserve margin is really a decision about how much reliability is economically desirable; since more reliability generally costs more, some estimate has to be made of when to stop. The criterion, which again is often only implicit, is that capacity should be expanded until the marginal cost of capacity and the expected marginal cost of short-

[13]The next three parts of this section (D, E, and F), on marginal capacity costs, identifying the peak period, and the assigning capacity costs to rating-periods, are based on the work of National Economic Research Associates.

age are equal. Depending on the characteristics of the system, this rule gives an implicit reserve margin. Different customers might be willing to tolerate different levels of reliability. Some industries would be prepared to take an interruptible service if the price were lower, while others would go so far as to build their own backup generators to further reduce the already small probability of a power failure. It would be possible to provide tariffs that distinguished kilowatts by their reliability. Anyhow, the cost of an appropriately determined reserve margin is properly part of the marginal capacity cost of the utility.

E. IDENTIFYING THE PEAK PERIOD

The question of identifying the peak period, and effectively tying prices based on marginal costs to peak demand, is perhaps one of the most subtle and confusing parts of any approach to peak-load pricing. Although after the fact a particular hour of the year can be identified in which the peak demand actually occurred, it is generally unknown before the fact just when it will come or just how high it will be. System planners do have some expectations about how high the peak load will be, as well as a good idea of what the potential peak hours will be. These are the same kinds of expectations they must use in designing the system, and providing for reserve capacity and maintenance scheduling. Since the utility usually has to establish a set of rating-periods and associated prices before the fact, knowing the potential peak hours and the expected costs is essential for establishing rating-periods and estimating marginal cost.

More technically, demand is uncertain for a particular period of time in the future at any set of prices. Normally a good deal is known about the probability distribution that characterizes the uncertain demand, but that is about the best that can be done ex ante. In many parts of the country, load does not vary greatly from one hour to the next during the daytime hours in a particular season. In such cases, companies often face fairly broad daily peaks where the load at one hour is little different from that in several other hours. Furthermore, the average daily peak-load for the potential peak days as a group may be very close to the actual peak. Such load characteristics are generally the easiest for which to define pricing periods and to closely tie together prices charged and marginal costs incurred.

Uncertainty about demand, given current prices, is not the only thing which one must take account of in establishing rating-periods and estimating the relevant marginal costs. The amount of available capacity itself varies, both because of scheduled maintenance and because of forced outages. Therefore in establishing rating-periods and assigning the appropriate capacity costs (in accordance with expected shortage costs), one must not only look at the expected load, but also compare the expected load with expected available capacity. Insofar as maintenance cannot be shifted around easily, additional consumption, even at periods that do not have the very highest expected peaks of the year, may still involve expected curtailment costs in the short run and require additional reserve capacity in the long run. Prices for such periods must reflect a share of the marginal-capacity cost.

Finally, in moving from the current system of pricing to a system of peak-load pricing, patterns of use are likely to change. Some people have concluded that the only effect of peak-load pricing will be to shift the peak to some other time, and that the utility will end up chasing peaks around. Much of this concern stems from a misunderstanding about how to apply the principles of peak-load pricing. These critics envision loading all the capacity costs on to a very small number of identifiable peak hours. If in fact the utility were to do this, if it were a question of making 2:00 p.m. on Wednesday 20 December (for example) the hour to which to assign all the annual marginal capacity cost, then indeed it would be reasonable to presume that much load would simply shift from 2:00 p.m. to 1:00 p.m. or 3:00 p.m. But it is not proposed to do anything of the sort. Even at current prices, the potential peak period is fairly wide, encompassing several daytime hours between 7:00 a.m. and 11:00 p.m. And by examining load-research data for different classes of customers, the utility could also get a very good qualitative feeling for where the largest increases in consumption were likely to come. It could then use this information to refine the definition of the potential peak periods. Capacity costs would then be reflected in the prices charged during a relatively large number of hours rather than a handful, and problems of shifting peak would be minimized.

Shifting peaks and uncertainty of demand present absolutely no conceptual problems. In theory, if the utility had estimates of all the demand elasticities, cross-elasticities, etc., it could calculate the marginal opportunity cost of consumption taken at any particular time. That it did not have all the information that it would like (and remember that econometric estimates by their very nature are uncertain and imply a probability distribution of behaviour) would not mean that progress could not be made. System planners can define fairly well where the potential peaks are now, and also have some fairly good feeling for the kinds of short-run and long-run responses that could result. Since most of the responses by far will occur over several years, the utility should learn more about the relevant demand functions as time goes on, and incorporate what it learned as rates were adjusted in the future.

It is important to remember that since the utility lacks perfect information, some judgement must be used. The initiation of peak-load pricing is a sequential process for which the utility must be prepared to make adjustments as more information becomes available to better inform judgement. In the effort to estimate the appropriate marginal costs for the Corporation, cautious assumptions have been made about demand responsiveness. While the total demand for electricity has been closely analysed in recent years, demands by time of day and season of the year have not received much consideration, largely because the relevant data were not available. Better information about own price and cross-elasticities is important for determining both pricing-periods and the revenue yield the utility would require from any particular set of rates.

The importance of 'learning by doing' in this pricing-scheme (something characteristic of all new production processes) means that the movement to peak-load pricing must be accompanied by an extensive program of load research. The information obtained from such a program would form the basis of any fine-tuning required over time.

F. ASSIGNING MARGINAL CAPACITY COSTS TO DIFFERENT RATE PERIODS

The problem of how to assign marginal capacity costs to different hours is closely related to the process by which rating-periods are selected. In choosing rating-periods, one should look not only at the height of the load at different hours, but also at the relationship between demand and available capacity. It is also important to estimate the probability that available capacity will fall short of demand during a particular set of hours. During peak periods this probability is likely to be fairly high; during one

or more shoulder periods it will be somewhat lower; and during off-peak periods it will approach zero. The short-run marginal cost of an additional kilowatt-hour is given by the expected marginal energy cost plus the expected shortage cost. One can therefore express the expected shortage cost in terms of capacity costs for a system with the optimal level of capacity, and so arrive at a rule for assigning an appropriate share of the marginal cost of capacity during each different rating-period according to the relative probabilities of shortage. One should also take account of differences in the costs of actual outages during different periods of the year.

The point is that one cannot attribute the marginal cost of capacity and reserve to only the one hour of peak demand. Planning-criteria based on reliability take into account the need for adequate capacity at all hours; and while the hour of peak demand is generally the hour of greatest exposure (although not invariably), other hours bear a risk related mainly, although not exclusively, to the level of demand. One can therefore think of the responsibility for capacity as being a graduated peak responsibility, rather than one attributable to a single hour, and use the relative value of loss-of-load probability (LOLP) in each hour to estimate the graduated responsibility for capacity costs.

The LOLP is derived by comparing the available capacity with the demands on it, including demands which are probabilistic rather than determinate. Starting with the load-duration curve, which gives a most probable level of demand for energy at each hour, the planner adds demand for planned maintenance; this gives a determinate total demand level.

On the supply side, each machine has a history of forced outage, from which one can estimate the probability of its being out for unplanned reasons. Looking at the system as a whole, one can arrive at a probability distribution which shows the (probable) amounts of forced outage on the system at any given level of demand. The level of demand itself may also be variable, particularly perhaps with the weather, and the probability distribution of demand can also be estimated. While all companies aim to have more capacity than the determinate peak demand projection, they have to estimate how many times they will have a conjunction of these probabilistic events. How often, for instance, will three plants be subject to random outage in the peak period? Or how often will freak weather be likely to make the temperature soar just when the largest machine is out for planned maintenance in weather which would normally be balmy? Probabilistic methods can enable the planner to answer these questions and calculate the loss-of-load probability at each hour for various possible levels of reserve.

These LOLP estimates reflect information on the probability distribution of demands over the year, maintenance schedules, and forced-outage rates. In making use of these probability estimates, the guiding principle is that the price at any period of time should reflect the expected marginal cost of energy plus the expected marginal shortage cost. The expected marginal shortage cost can be expressed in terms of the marginal cost of capacity, thus allocating the marginal cost of capacity to all periods which have a significant shortage probability, in accordance with the relative probabilities in the different rating-periods.

Further discussion of the theoretical and applied aspects of time-of-use pricing is found in Volume VII *Costing-Methodology for Determining Marginal Costs*. In particular, Section II on the choice of Costing-Periods and Appendices I and II are useful supplements to this Section. In addition Volume VIII, *Detailed Proposals for Rate Structure*, illustrates the application of the theory and costs to time-of-use rates.

V. MARGINAL-COST PRICING AND THE REVENUE REQUIREMENT

A. THE PROBLEM

Prices based on marginal costs could yield more revenue than the utility requires. This is because the revenue requirement is based on historical accounting-costs. Because the revenues generated by the rate structure must neither exceed nor fall short of the revenue requirement, a potential problem exists: Is it possible to reconcile the objective of efficiency with meeting the constraint of the revenue requirement?

This revenue surplus seems to be derived from three main sources. *First, although the electric utility industry is undoubtedly characterized by economies of scale, recent econometric studies show that most large firms have reached the point where there are few further economies of scale they can make. Second, the effect of environmental controls has been to increase the costs of new plant. Third, the effect of price increases for capital, labour, and materials for new plants has been to raise marginal costs faster than the historical-cost rate base.*

The effect of inflation is clearly to raise the money cost of plants from year to year. This would not in itself mean that marginal costs would exceed revenue requirements, except that depreciation policies have been fundamentally mistaken in an economic sense.

Depreciation should be the contribution the users make in a given year for the use of a machine in the year, and should therefore reflect the change in the assets value over the year. When technology is moving quickly and new improvements reduce the cost of replacements, the value of the asset will quickly decline. If, on the other hand, prices of new equipment are rising, then economic depreciation may in fact be negative: the economic value of a machine may actually rise in a particular year. If the economic value were correctly stated on the books, the gross return on the net book value plus the variable cost of operating the old plant would produce a cost of service exactly equal to that of a new plant. This would then eliminate most of the revenue gap.

In periods when inflation is pushing the reproduction costs beyond the historical cost, and when old plant is nonetheless depreciated on the books by straight-line methods, marginal costs are likely to exceed revenue requirements based on original cost. The revenue requirements are based not only on current spending for such items as fuel and labour, but also on those depreciation schedules which overestimate the loss in value early in the life of the plant. The resulting valuation of the rate base on which return is earned is a mixture of variously depreciated properties bearing no relation to current value. It should surprise no-one that in periods of inflation revenue requirements will almost never equal marginal costs, and will generally be below them. The reverse is true in periods of technical progress.

Inflation has one further effect. The bonds which were sold at 3 per cent when there was no inflation are now holding down revenue requirements, because interest rates have since risen to include a premium for inflation. Marginal debt prices are above average historical debt prices. The same is not true of equity capital for privately owned utilities, since the regulatory process generally permits the return on old equity to equal the rate of return on new equity. These, then, are the sources of the gap. What is the best way to meet it?

The classic solution is to require the utility to price all units of its commodity at marginal cost and leave the government tax away the surplus revenues or, alternatively, subsidize any deficit that results from marginal-cost pricing. However, there are three 'second-best' pricing-methodologies which let the utility just meet its revenue requirement while maintaining the objective of efficiency. They are:

1. Multi-Part Pricing,
2. Block-Rate Pricing, and
3. The Inverse-Elasticity Rule.

B. MULTI-PART PRICES AND BLOCK RATES

Multi-part prices are well-known devices to improve the efficiency of public pricing where a revenue constraint exists. Indeed, when pure marginal-cost pricing leads to surplus revenues, multi-part pricing and block-rate pricing let the electric utility meet the revenue requirement without imposing use and output distortions on the economy. Both play on the fact that the necessary and sufficient condition for efficiency pricing is that the price for the last unit sold must equal marginal cost, and that markets must clear. Both the simple rule that the price of each unit of output should be set at marginal cost and the more general rule that the price for the last unit sold must equal marginal cost lead to identical decisions about use and production. The difference is that the utility can adjust the prices for intramarginal units of consumption and still retain efficiency. That is, if the customer's demand function is known, the utility can charge prices which deviate from marginal cost for intramarginal units, reflecting the use value of each additional unit of service. There are limits to the deviations of price from marginal cost. For example, when there is a deficit, prices will rise above marginal cost for intramarginal units up to the point where the use value of an additional unit is less than the marginal cost to the consumer. The utility then returns (obtains) extra revenue by increasing (extracting some) consumer surplus of individuals through lower (higher) prices on intramarginal units.

Pricing-methodologies of this type may take two basic forms. First, when there were surplus revenues, the utility could make a credit to the charge for becoming and remaining a customer, and then proceed to charge marginal capacity and operating-costs according to use. The conditions are that the customer charge credit must not exceed the consumer surplus realized from efficient marginal-cost pricing, and that (ignoring income effects) decisions about use must not be affected. An alternative is to price some or all intramarginal units at their 'value' to customers, with a charge equal to marginal costs for the last unit of output sold. When there was a revenue deficit, such a rate structure would look like a declining block with the end rate based on marginal costs. When there was a revenue surplus, it would appear as an inverted block with the end rate based on marginal costs.

Multi-part pricing may be considered in simpler terms. If all characteristics of the electric utility's output were priced at marginal cost (customer, demand, peak energy, and off-peak energy), then the revenues generated might exceed or fall short of the revenue requirement. That is because the revenue requirement is based on historical accounting-costs.

It is necessary, then, to depart from marginal cost pricing where the cost information that the price signal provides is least important to the user's decision. Of the decision components outlined, the customer's decision to join the system is least sensitive to price. The other decision components (use, rate of use, and time of use) are much more sensitive to price. It is important, then, to maintain the integrity of the cost information in the

price about use, rate of use, and time of use. On the other hand, the information relayed through the customer charge may be distorted over a reasonable price range. Hence, the demand charge and energy charge should reflect their marginal cost. The customer charge may be adjusted upwards (to recover a deficit) or downwards (to eliminate a surplus) with minimal distortion to price information and consumers' decisions.

The method of returning the surplus here must be chosen carefully, because it may affect the distribution of wealth in society. Hence it is important to define some criterion for the distributional value-judgement to be reflected in returning surplus revenues. What this value-judgement is to be is clearly not a question of economics, but rather of society's preferences. However, once it has been defined, economics can help one to arrive at an efficient pricing-methodology that incorporates the distributional constraint.[14]

The multi-part rate is recommended for use in the detailed rate-structure design of Volume VIII. The method of returning the surplus is discussed in some detail there.

A multi-part rate structure is theoretically equivalent to a block-rate structure. Of the two, however, a multi-part rate is preferable, because it requires fewer assumptions and thus poses fewer problems. There are several informational difficulties to implementing either a multi-part rate or a block rate. For both, the utility needs some minimum estimate of the surplus customers would realize from straight marginal-cost pricing. This is hard to estimate, and would vary widely among customers, even those classified as residential. Furthermore, in applying multi-part pricing one must take care lest it should distort users' choices among alternative energy sources, because the customer charge may lock out potential customers. This will probably not apply to an electric utility, since almost any enterprise needs to take electricity, whether it uses it as a primary production input or not.

However, using block rates for intramarginal units of electricity consumption while preserving the proper pricing relationship on the margin poses greater difficulties. Even if all customers were identical, ideal block rates would require detailed knowledge of consumer preferences such as is not available at present. More important is that customers are far from identical. Hence the objective of efficiency is not met. If a large group of customers finds one of the early blocks relevant for their use, while only the end block is based on marginal cost, then the objective of efficiency is lost.

There are some final observations which should be made on block-rate structures. It has been claimed in the past that a declining-block rate tracks costs; this is simply not true.

It is generally understood that large volumes of consumption tend, on the average, to involve lower average costs than smaller volumes, because distributing costs tend to be fixed on a per-customer basis. But it does not follow that failure to retain a declining-block rate would produce a structure out of line with cost, at least not as a matter of economics. Declining average costs do not justify declining marginal rates that is, declining rates for added units of consumption. The tendency for customer costs to be fixed calls for a flat charge; it does not justify declining charges per kWh of consumption. That customer costs are more or less fixed is no reason for inferring that marginal costs are different for larger users than for smaller ones.

Another justification sometimes offered for the declining-block rate structure (but one that is even less well founded) is the possibility of further economies of scale. The fact that marginal costs may be below average does not in itself justify confronting high-volume customers with a lower marginal charge than low-volume customers. It is only insofar as, by chance, use in some of the larger-volume blocks may be more elastic than in the smaller volume ones that, given the impossibility of marginal-cost pricing across the board because of the revenue requirement, this discriminatory system may prove to be second best.

In the early years of the electric utility industry, when the primary objective of utilities was electrification, there was some justification for a declining-block rate. The argument was that under a multi-part rate, a high customer charge could deny electricity to some who could have afforded it under a declining-block rate structure.

C. THE INVERSE-ELASTICITY RULE

There is another price structure which in theory generates just enough revenue meet the requirement and still preserves the objective of efficiency. It is a second-best approach to pricing, which a utility can avail itself of when it faces a revenue constraint. It has come to be known as the pricing-approach of the 'inverse-elasticity rule'.

This approach to pricing requires the utility to discriminate among customer classes. That is, identical prices are charged for each unit of output within a customer class, but the price per unit of output will be different for different classes as their demand characteristics differ. The classic statement of this result is that a utility subject to the constraint of a revenue requirement should meet it by setting prices in the separable markets (where demand functions are not interdependent) so that the percentage deviation of price from marginal cost in each market is inversely proportional to the elasticity of demand in that market. Other things being equal, in a case of surplus revenues, the set of quasi-efficient prices would be lowest in the least price-elastic market and highest (closest to long-run marginal cost) in the most price-elastic market.

Hence where marginal cost exceeds the average cost of output, the utility should set all its prices below the respective marginal costs. But in those markets where the demand is most elastic, the price should come closest to marginal cost, so as to minimize encouraging growth in demand at rates that do not cover the value of the resources consumed in meeting it. In markets where the demand is less elastic, the rates should correspondingly be that much lower than marginal costs, because in these markets lower prices will least encourage uneconomic consumption.

One can also think of this pricing-method as a two-phase setting of rates. First, all rates are set to equal marginal cost. This would generate, in these circumstances, excess revenues. Then the surplus revenues are distributed among the customers as rate 'reductions'. In order not to encourage significant additional consumption, most of the 'reduction' should go to the customers with relatively inelastic demand, while only small 'reductions' should go to customers with elastic demands. Thus, the prices will tend to minimize encouraging avoidable uneconomic growth in the use of electricity. 'Uneconomic' growth means growth beyond what the consumer would be willing to use if he had to pay a price equal to the marginal costs of production. This way of re-

[14]See for example J. Green, "Two Models of Optimal Pricing and Taxation", *Oxford Economic Papers* (1975).

turning surplus revenues to customers produces a pattern of use among customers which is as close as possible to what would prevail under marginal-cost pricing.

The fundamental weakness in this method is that, at present, there is simply not enough information about the price elasticity of demand for different customer classes. Moreover, it is discriminatory in the sense that prices reflect the market characteristics of the customer class, and not the cost characteristics of the commodity.

VI. ARGUMENTS AGAINST MARGINAL-COST PRICING

Considerable debate has surrounded marginal-cost pricing over the years. It is therefore useful to consider here the arguments for rejecting it. These have shown considerable variety both in character and weight, and only the most influential are summarized and reviewed here. The three main arguments against marginal-cost pricing are

1. The 'Second-Best' or 'All-or-Nothing' Argument,
2. The Wealth-Distribution Argument, and
3. The Calculation-of-Marginal-Cost Argument.

The last section outlines the problems of carrying the theory into practice. Most of these have been identified in earlier chapters, and are merely summarized for convenience.

A. THE 'SECOND-BEST' OR 'ALL-OR-NOTHING' ARGUMENT

The argument used most often against marginal-cost pricing is the second best position. It has been emphasized that the rule of marginal-cost pricing is essentially an all-or-nothing rule, in the sense that there is no advantage in applying it to one industry or service if it is not simultaneously applied to all others. Indeed, by applying it to one industry when it is not applied elsewhere, one might well be doing havoc rather than good, so far as the effect on the allocation of resources is concerned.

A facile answer might be that what one needs to do then is to make price bear a relation to marginal cost equal to its average relation in all other industries. But even if one could find out what that average was, only given very special assumptions (for example about the transferability of resources) would this modified version of the rule retain any validity. Some critics have gone so far as to assert that in such a situation one cannot set any general rule whatsoever, and that whenever a large sector of industry is non-optimal, there is no possible means of knowing in which direction a second-best solution will lie.

However, this criticism is not as devastating as it may appear. As E.J. Mishan has noted, "We cannot sit by and sadly suck our thumbs under the sign of second best". For in the first place, the pricing-objective of any public utility should be efficiency. That is, the price structure should contribute to the efficient allocation of the resources devoted to producing the utility's output. The utility itself is not concerned or empowered to deal with the more general issue of allocating all society's resources efficiently.

The second-best theorem really contrasts the different results which arise from a 'partial-equilibrium' as opposed to a 'general-equilibrium' approach to policy. Thus, it may be argued that since everything depends on everything else, adjusting any one price will entail adjusting all others.

However, everything does not depend on everything else in any significant degree. There is evidence to suggest that interdependece between activities is not as complete as the second-best theorem would seem to imply. The basic concept is that there are "key" sectors or activities which play a crucial role in the economy. What is relevant to marginal pricing is the relative price relationship of the commodity (sector) under consideration to substitute and complementary goods. This point has been clearly stated by I.M.D. Little, in *The Price of Fuel:*

> *The importance of having relative prices and relative marginal costs equal depends clearly on substitutability ... if there were zero substitutability it would not matter what the ratio was. Since fuels are far better substitutes for each other than any of them is for food, medicine or education, a change which gets relative prices and costs equal for the different fuels at the expense of greater inequality of price and cost between food and fuel may be reasonably judged a good one.*

Only if prices in the substitute markets deviate from marginal costs, and if the government is powerless to adjust these prices, may deviations of price from marginal cost be called for to preserve efficiency in the market under consideration. However, in taking such considerations into account in pricing, the starting-point must still be marginal cost. Decisions to deviate from marginal cost in one direction or another can only be made intelligently if one knows what the marginal cost actually is.

The accompanying relatively simple formula may be used to try to evaluate whether a particular price change will lead to an improved allocation of resources.[15]

The first term in the expression above represents a measure of the change in economic efficiency in the electricity market itself. The second term provides a "second-best" correction by summing the second-best effects over all commodities (i) whose demand is affected by the price of electricity.

For example, if all other commodities are priced at marginal cost, the second term disappears. If the original price were less than marginal cost and were changed to a price equal to marginal cost, the value of the first expression would be positive (a negative times a negative). If, on the other hand, there were some commodity (i) whose price was below marginal cost and which was a substitute for electricity, the effect of increasing the price of electricity would be to increase consumption of that commodity. The second term would now be negative, and the sign of the entire net-benefit equation ambiguous. One would need more detailed information on the relative sizes of the price responses in the electricity market and its complement market and the size of the deviations of price from marginal cost to come to a definitive conclusion. Other situations involving complements, or prices greater than marginal cost, can be analysed in the same way.

Primary practical concern about second-best prices has revolved around the relationship between the prices of oil and natural gas and their marginal costs. With empirical information about the price-responsiveness of consumption of oil and natural gas to electricity prices, and the relationship of price to marginal cost, one could use the net-benefit equation to learn whether a particular movement increased economic efficiency, and also to search for a price for electricity that maximized the net-benefit equation.

In conclusion, second-best considerations may complicate pricing-policy, but they do not make rational pricing impossible. By using the net-benefit relationship above, one should be able to get a fairly good feeling for whether a price change will make things better or worse, and also some help at reaching an optimal second-best price for electricity that accounts for distortions elsewhere.

B. THE WEALTH-DISTRIBUTION ARGUMENT

It has been pointed out many times that one need accept the results of marginal-cost pricing as efficient only if one is willing to place a similar evaluation on the distribution of wealth. Two positions have to be considered: the pre-existing distribution of wealth, which decides how many dollar votes each buyer has in

[15]See Ralph Turvey, "Price Changes and Improved Resource Allocation", *Economic Journal" 84, No. 336 (December 1974)*

$$\text{Net Benefit} = (Q_1^e - Q_o^e)\ \frac{(P_o^e - P_1^e)}{2}\ - \ MC_e + i\Sigma(Q_1^i - Q_o^i)\ (P_i - MC_i)$$

Where: Q_o^e = the quantity of electricity consumed at the old price (P_o^e).

Q_1^e = the quantity of electricity consumed at the new price (P_1^e).

Q_o^i = the quantity of some other commodity whose consumption depends on the price of electricity, at the initial price of electricity.

Q_1^i = the quantity of this other commodity at the new price of electricity.

P_i = the prevailing price of the other commodity i.

MC_i = the marginal cost of the other commodity i.

deciding what to order the economy to produce, and the distribution of wealth that results from equating price to marginal cost. It should be noted that this is an ethical argument, not an economic one.

The pricing-rule does suggest that if one wishes to redistribute wealth, one might do it better through lump-sum taxes (for example, on rents, income, or inheritance) and money transfers (for example, a negative income tax) than by departing from the requirements of economically efficient pricing. If customers are to make the best choices, then the necessary and sufficient condition is for marginal use of a good or service to be priced at its marginal cost, however wealth may be distributed or redistributed.

C. THE CALCULATION-OF-MARGINAL-COST ARGUMENT

Another major objection to marginal cost as a basis for pricing is that it has no unique definition, but rather a considerable variety of definitions, each contingent on the special circumstances of a particular case. It is accordingly an unsuitable basis for a general rule and impracticable to carry out. As W.A. Lewis claims,

> *There is no such quantity as the marginal cost of output; there is not even a simple choice between two quantities, short- and long-run cost; there is a large variety of costs to choose from, depending merely on how far ahead you choose to look; and this collection of costs itself varies from day to day as current commitments alter.*[16]

Such an objection is based on a misunderstanding of how the marginal rule works; for it ought to be regarded essentially as a rule not for fixing prices but for fixing *output*. All that is needed is to tell managers, when making decisions about the scale of output, to use that definition of marginal cost which fits the particular decision they are making. What they should do in any particular case is simply calculate the *additional* cost involved in the change or activity under consideration, and compare this with the value of the additional output. If the latter exceeds the former, the increase of output should be undertaken; but otherwise not.

Evidently, an assumption has slipped into this way of formulating the problem: namely, that the price is somehow independently given, whoever decides the output accepts it as such. This is simply not true: the marginal costs used for pricing are the same costs the system planner uses in planning additions to the system. It should be noted, though, that a trade-off of sorts is made between price stability and the theoretical ideal of "reactive marginal-cost pricing" based on short-run marginal costs proposed by William Vickrey[17]
Such a responsive or reactive price system would allow for continuous and instantaneous price changes in response to changing cost conditions.

D. OTHER ARGUMENTS

Several other arguments are levelled against marginal-cost pricing. Three of the ones most often cited are the following:

1. Social Costs

Critics have claimed, quite reasonably, that marginal net external costs must be incorporated into the price if the customer is to face the cost consequences of his decisions about use.

2. Customers' Decision-Making Abilities

Some critics maintain customers cannot make sensible decisions about the mix of goods and services they ought to buy, and that therefore such matters as how much electricity an individual uses ought to be decided by the utility, or by the government.

3. Revenue-Requirement Constraint

Other critics claim that it is impossible to meet the objective of the constraints of efficiency because of the revenue requirement.

The first two arguments were dealt with in Section III, while the third was the subject of Section V. It would be superfluous to repeat the arguments here.

---

[16]W.A. Lewis, *Overhead Costs* (London, 1949), p. 12.
[17]W. Vickrey, "Responsive Pricing of Public Utility Services", *Bell Journal of Economics and Management Sciences* 1971, pp. 337-346. As Vickrey himself notes, "Indeed the main difficulty with responsive pricing is likely to be not mechanical or economic, but political. The medieval notion of the just price as an ethical norm, with its implication that the price of a commodity or service that is nominally in some sense the same should not vary according to the circumstances, has a strong appeal even today."

APPENDIX I: Efficiency and Marginal-Cost Pricing: A Mathematical Demonstration

It is assumed here that the aim of any public enterprise should be to maximize social benefit minus social cost.

The difference between social benefit and social cost may be referred to as social surplus. Hence the aim of any public enterprise is to maximize social surplus. This objective is met by the economic pricing-rule of marginal-cost pricing.

In order to show the derivation of this pricing-rule, a formal statement of an objective function is made and maximized. Let P = p(Q), where P is society's willingness to pay for varying amounts of Q, or output of a particular commodity, and p(Q) is the willingness-to-pay function or the demand function. It is assumed that p(Q) is continuous and has continuous first-order and second-order derivatives.

The demand function, in this context, may be viewed as the marginal benefit function of society. Hence total benefits (B) are defined as the integral of the demand function. Thus $B = \int PdQ$. Then let C = c(Q) where C is the total social cost to society necessary to produce varying amounts of Q, and c(Q) is the total social cost function. It is assumed that c(Q) is continuous and has continuous first-order and second-order derivatives.

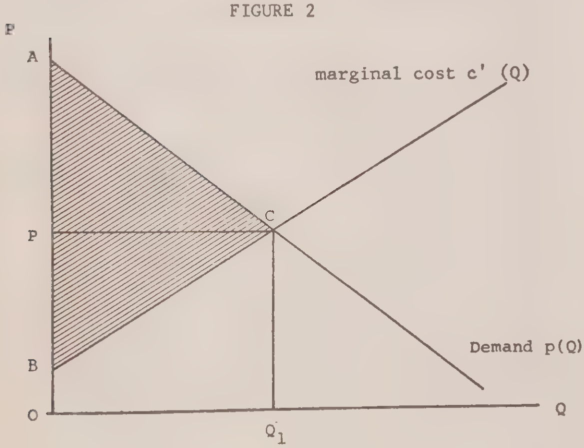
Surplus (S) is defined as the difference between what society is willing to pay for total benefits and the total social costs of producing the output. The surplus function can be defined as S = B - C, which becomes $S = \int PdQ - c(Q)$.

Surplus viewed in this way may be considered an objective function. Maximizing surplus yields the following condition of setting price equal to marginal cost: ds Old Q = P-(dC Old Q) = P - c'Q = 0; therefore P = c'(Q).

Therefore surplus is maximized at that level of output where price is equal to marginal cost.

This result may be shown geometrically as in Figure 2.

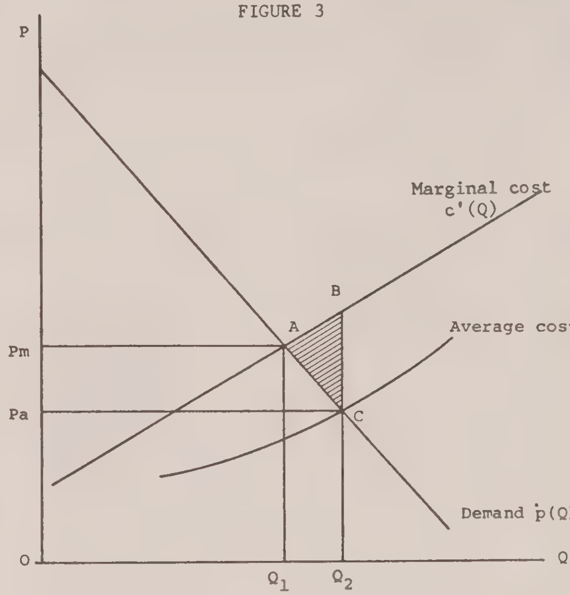
FIGURE 2



Surplus ABC is at a maximum at that level of output where price (P) is equal to the marginal cost of production at that level of output.

Consider the case where the level of output is greater than under marginal-cost pricing. That happens when price is set equal to average cost and the marginal cost of production is above the average cost. The relevant question is, Would such an increase in output lead to an increase or decrease in surplus? This case is illustrated in Figure 3.

FIGURE 3



The additional benefit derived from the additional consumption Q1Q2 is equal to the area Q1Q2CA. However, the additional costs of Q1Q2 units of output are equal to the area Q1Q2BA. That is, there is a net cost to society equal to area ABC by increasing the level of output from Q1 (under marginal-cost pricing) to Q2 (under average-cost pricing).

More formally, the problem can be assessed by comparing the total surplus at output Q1 under marginal-cost pricing and the total surplus at output Q2 under average-cost pricing. That is:

$$S(Q_2) - S(Q_1) = \int_{Q_1}^{Q_2} \left[p(Q) - c'(Q) \right] dQ$$

From the above analysis, more complex models may be developed showing efficient pricing-rules when the public enterprise is subject to a revenue-requirement constraint, a peak-load problem etc.

APPENDIX II: Efficiency and Fairness in Pricing

This volume has focused on fully defining the objective of efficiency for pricing and its implications in terms of marginal-cost pricing. Traditionally, however, 'fairness' has been cited as one of the primary criteria in rate-setting for electric utilities. It is often suggested that the objective of efficiency is in basic conflict with that of fairness objective. The problem in answering this charge is that fairness is seldom defined. The purpose of this appendix is to try to remove some of the veils of mystery from the so-called criterion of fairness.

A. THE TRADITIONAL MEANING OF FAIRNESS IN ELECTRIC UTILITIES' RATE-MAKING

Traditionally, the fairness criterion has had two dimensions in the rate-making of electric utilities. The primary characteristic of fairness is a relevant cost-based rate that passes on the historical benefits of past investment decisions.[18]
The second characteristic of fairness is sometimes, mistakenly, interpreted as a corollary of the first. This second characteristic requires distributing both joint and fixed costs fairly among customer classes and customers within classes. Traditionally, the costs which are fairly allocated among customers are historical accounting-costs. Because there are several methods for allocating these costs among customers, judgement must be used in establishing and/or choosing a fair cost allocation. However, it is contended that reasonable people who are knowledgeable about the techniques of production can establish acceptable guidelines for allocating fixed and joint accounting-costs to customers fairly.

It is important to note that the fairness criterion does not deal in any way with socio-ethical objectives for distributing wealth. Such social objectives are properly the business of government.[19]

B. THE FAIRNESS CRITERION IN RATE-MAKING: AN ELABORATION

Because fairness in rate-making has never been rigorously defined in the traditional rate literature, its meaning and application tend to vary from one rate-maker to another. Yet in its simplest form, the fairness criterion has great appeal to rate-makers and the public alike. However, if it is to be a meaningful criterion for setting rates, some attempt must be made to define the term more rigorously than in the past. This task is not as intimidating as it first may appear.

Historically, certain criteria have been commonly appealed to in choosing, from among the countless alternatives, a particular pattern of government services and the means of financing. Most of these criteria have become government economic goals (e.g., full employment, economic growth allowing for environmental concerns, and price stability).

The fairness or equity criteria, however, have been viewed as characteristics which the community desires all public finance instruments (taxes, government expenditures, prices of public utilities, etc.) to possess, rather than as goals to be achieved by the use of these instruments.

On some fairness criteria the community will be found united, once the question at hand is well understood. These criteria raise no conflict-of-interest issues; they may be termed the agreed criteria of fairness. The traditional view of fairness in rate-making deals with agreed criteria. An example would be reducing uncertainty in the utility's prices.

Other equity criteria provoke sharp differences of opinion, since they call for making some persons worse off in order to benefit others. They may be called conflict-of-interest criteria. A conflict criterion, therefore, supplies a standard, a guide to policy, but one which is imposed against the wishes or judgement of some members of the community. An example of this would be to use utility prices to redistribute wealth (for example, through lifeline rates). The traditional view of fairness in rate-making does not concern itself with conflict criteria. Utilities have traditionally viewed conflict criteria as lying in the domain of government. Utilities have not viewed their rate structure as policy instruments to deal with social objectives involving conflict criteria.

C. THE CONSENSUS CRITERIA OF FAIRNESS IN ELECTRIC UTILITY RATE-MAKING

There is a generally accepted standard of fairness for all public finance measures: equal treatment of those equally circumstanced. It is "a principle predominantly founded in analogy with equal treatment before the law".[20] A corollary is almost equal treatment of those almost equally circumstanced. Only relevant circumstances are to be considered, of course: relevant, that is, by community consensus. If there is a conflict of opinion about the relevance of a particular circumstance, then the issue falls into the class of conflict criteria.

Equal treatment of those equally circumstanced, and almost equal treatment of those almost equally circumstanced, have six implications for the price structure of an electric utility, as follows:

1. The price-structure must maintain the integrity of the concept of cost pooling
2. There should no seniority rights in price-structure. All consumption is always new, for the customer may decide to discontinue it at any moment.
3. The price structure should be impartial. There should be no undue discrimination; all end users should be priced at the same rate.
4. The price-structure and changes in the price level should be defined clearly, so that the customer knows the price he will pay if he takes a specific course of action. This is the criterion of certainty in prices.
5. Changes in Corporate policy (for example, level of system reliability) should not lead to abrupt changes in the quality of service received or price charged. This is the criterion of continuity in prices.
6. In a time of increasing costs, customers should receive the economic rent known as the historical benefits of investment in a way that does not distort the distribution of wealth. That is, the price structure for the electric utility should be distributionally neutral; it should not be used for social engineering. Similarly, when costs are decreasing customers should receive the negative economic rent known as the historical burden of investment in a way that does not distort

[18] In a decreasing-cost situation, fairness implies a relevant cost-based rate that passes on the *burden* of historical investment decisions.
[19] See for example, "Public Power Rate Policies In the Next Five Years", by J.B. MacDonald, Manager, Power Market Analysis, Ontario Hydro Corporation, in *Electric Rates: Current Practices and Problems* , (published by the American Public Power Association, May 1975). Mr MacDonald states the position of fairness and social objectives quite clearly: "Departure from a fair distribution of cost for social purposes should be instituted by government action only" (p. 7).
[20] Douglas Dosser, "Economic Analysis of Tax Harmonization", vol. I, p. 20 in *Fiscal Harmonization in Common Markets*, ed. Carl S. Shoup (New York, 1967)

the distribution of wealth. This is known, then, as the criterion of distributional neutrality.

D. THE CONSENSUS CRITERIA OF FAIRNESS AND THE EFFICIENCY OBJECTIVE

It should be clear that there is nothing necessarily incompatible between the foregoing agreed criteria of fairness based on social preferences and the objective of efficiency for the pricing-structure of an electric utility. More specifically, marginal-cost pricing and fairness are not incompatible.

In the past, traditional rate-makers have nearly always assumed that a fully distributed average-cost (FDC) pricing-system could fairly return the historical benefits of investment to customers. However, such a price structure does not meet the objective of efficiency.

On the other hand, it was felt that marginal-cost pricing would not fairly return the historical benefits of investment. This is not necessarily so. The surplus revenues marginal-cost pricing yields constitute an economic quasi-rent derived from historical investment. The fairness criterion of distributional neutrality requires returning this economic rent to the customers in a way that minimizes distortions to the current distribution of wealth. This criterion acts as a constraint, then, in designing a rate-structure which meets the objective of efficiency. The recommended method of returning the surplus revenues to large users attempts to do so. In operational terms, it comes as near distributional neutrality as is feasible at this time. Ultimately, the method of surplus return or distribution of historical benefits of investment now proposed for large users should be extended to all customers. This is because it recognizes that the marginal utility of money may differ among customers: that is, an extra dollar of income may be worth more to a poor man than to a rich man. The pricing-rule not only allows for this, but allows for it with each individual customer, thus ensuring distributional neutrality.

APPENDIX III: A Select Bibliography of Marginal-Cost Pricing

Acknowledgement is given to our consultants, National Economic Research Associates of New York, whose permission to draw on their work has expedited completion of this volume. The following select bibliography provides a guide to the more recent literature on marginal-cost pricing. It is not meant to be exhaustive.

1. Anderson, D. "Models For Determining Least-Cost Investments in Electricity Supply", *The Bell Journal of Economics and Management Science* (Spring 1972), pp. 267-299.
2. Bailey, E.E. *Economic Theory of Regulatory Constraint*, Toronto, 1973.
3. Bailey, E.E. "Peak Load Pricing under Regulatory Constraint", *Journal of Political Economy* (1972), pp. 662-79.
4. Bailey, E.E., and L.J. White. "Reversals in Peak and Off-Peak Prices", *The Bell Journal of Economics and Management Science* (Spring 1974), pp. 75-92.
5. Balasko, Y. "Optimal Forms of Electricity Tariffs", UNIPEDE, April 1975.
6. Baumol, W.J., et al. "Costs and Rail Charges", pp. 117-129 in *Transport*, ed. D. Munby, 1968.
7. Baumol, W.J. and D.F. Bradford. "Optimal Departures from Marginal Cost Pricing", *American Economic Review* (1970), pp. 265-83.
8. Bergson, A. "Optimal Pricing for a Public Enterprise", *Quarterly Journal of Economcis*, November 1972.
9. Berlin, E., C.J. Cicchetti, and W.J. Gillen. *Perspective on Power*, Cambridge, Mass., 1974.
10. Bonbright, J.C. *Principles of Public Utility Rates*, New York, 1971.
11. Boiteaux, M. "On the Management of Public Monopolies Subject to Regulatory Constraints", *Journal of Economic Theory*, September 1971.
12. Brown, G. (Jr), and M.B. Johnson, "Public Utility Pricing and Output Under Risk", *American Economic Review*, March 1969.
13. Cicchetti, C.M. and J. Jurewitz. *Studies in Electric Utility Regulation*, Cambridge Mass., 1975.
14. Crew, MA., and P.R. Kleindorfer. "Peak-Load Pricing With a Diverse Technology", *The Bell Journal of Economics and Management Science*, Spring 1976.
15. Crew, M.A., and P.R. Kleindorfer. "Marshall and Turvey on Peak Load or Joint Product Pricing" *Journal of Political Economy* (1971), pp. 1369-77.
16. Crew, M.A. and P.R. Kleindorfer. "On Off-Peak Pricing. An Alternative Technological Solution" *Kyklos* (1975), pp. 80-93.
17. Crew, M.A., and P.R. Kleindorfer. "Optimal Plant Mix in Peak Load Pricing" *The Scottish Journal of Political Economy* (1975), pp. 271-91.
18. Cross, J.G. "Incentive Pricing and Utility Regulation", *Quarterly Journal of Economics*, May 1970.
19. Davidson, R.K., *Price Discrimination in Selling Gas and Electricity*, Baltimore, 1955.
20. Dixon, P. "The Costs of Average Cost Pricing', *Journal of Public Economics* (1972), pp. 245-56.
21. Dreze, J.A. "Some Postwar Contributions of French Economists to Theory and Public Policy", *American Economic Review*, (June 1964 Supp.), pp. 1-64.
22. Feldstein, M.S. "Equity and Efficiency in Public Sector Pricing: The Optimal Two-Part Tariff", *Quarterly Journal of Economics* (1972).
23. Feldstein, M.S. "Distribution Equity and the Optimal Structure of Public Prices', *American Economic Review* (1972).
24. Foster, E. and H. Sonnenschien. "Price Distortion and Economic Welfare" *Econometrica* (1970), pp. 281-97.
25. Green, H.A.J. "Two Models of Optimal Pricing and Taxation" *Oxford Economic Papers* (1975).
26. Joskow, P. "Contributions to the Theory of Marginal Cost Pricing", *Bell Journal of Economics*, Spring 1976.
27. Joskow, P. "Pricing Decisions of Regulated Firms: A Behavioral Approach", *Bell Journal of Economics and Management Science*, Spring 1973.
28. Kahn, A.E. *The Economics of Regulation*, 2 volumes, New York, 1971.
29. Kolm, S.C. "Footnotes to Marcel Boiteux's Value - Constrained Second Best", *Journal of Economic Theory* (1971).
30. Kolsen, H.M. "The Economics of Electricity Pricing in N.S.W.", *Economic Record*, December 1966.
31. Leibenstein, H. "Allocative Efficiency vs. x-efficiency", *American Economic Review*, June 1966.
32. Lipsey, R.G. and K. Lancaster. "The General Theory of Second Best", *Review of Economic Studies* (1956), pp. 11-32.
33. Little, I.M.D. *The Price of Fuel*, Oxford University Press, 1953.
34. Littlechild, S.C. "Peakload Pricing of Telephone Calls", *Bell Journal of Economics and Management Science* (Autumn).
35. Meek, R.L. "The New Bulk Supply Tariff For Electricity" *Economic Journal* (1968).
36. Meek, R.L. "An Application of Marginal Cost Pricing: The 'Green Tariff' in Theory and Practice", *Journal of Industrial Economics*, July and November 1963.
37. Mohring, H. "The Peak Load Problem with Increasing Returns and Pricing Constraints", *AER*, September 1970.
38. Moore, T.G. "The Effectiveness of Regulation of Electric Utility Prices", *Southern Economic Journal*, April 1970.
39. Nelson, J.R. (ed.) *Marginal Cost Pricing in Practice*, Eaglewood Cliffs, N.J. 1964.
40. NERA (National Economic Research Associates), *Proceeding As To Rate Design For Electric Corporations: Testimony before the New York Public Service Commission, Case Number 26806*, Albany, New York, 1975.
41. Ng, Y. and M. Weisser. "Optimal Pricing With a Budget Constraint - The Case of the Two Part Tariff" *Review of Economic Studies* (1974).

42. Nguyen, D.T. "The Problems of Peak Loads and Inventories" *Bell Journal of Economics*, Spring 1976.

43. Oi, W.J. "A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly", *Quarterly Journal of Economics*, February 1971.

44. Phillips, A., and O.E. Williamson (eds). *Prices: Issues in Theory, Practice and Public Policy*, Philadelphia, 1968.

45. Pressman, I. "Peak Load Pricing", *Bell Journal of Economics and Management Science*, Spring 1971.

46. Rees, R. "Second Best Rules for Public Enterprise Pricing", *Economica*, August 1968.

47. Ruggles, N. "Recent Developments in the Theory of Marginal Cost Pricing", *Review of Economic Studies* (1949-50), pp. 107-126.

48. Steiner, P.O. "Peak Loads and Efficient Pricing", *Quarterly Journal of Economics*, November 1957.

49. Trebing, H.M. (ed.) *Essays on Public Utility Pricing and Regulation*, East Lansing, Michigan, 1971.

50. Turvey, R. "Peak-Load Pricing", *Journal of Political Economy (1968), pp. 107-113.*

51. Turvey, R. (ed.) *Public Enterprise*, Penquin Books, 1968.

52. Turvey, R. *Optimal Pricing and Investment in Selectricity Supply*, Cambridge Mass., 1968.

53. Turvey, R. *Economic Analysis and Public Enterprises*, Totawa, N.J., 1971.

54. Turvey, R. "Marginal Cost" *The Economic Journal* (1969).

55. Turvey, R. *Testimony before the New York P.S.C., Case Number 26887*, Albany, New York, 1975.

56. Turvey, R. "Price Changes and Improved Resource Allocation" *Economic Journal* (1974).

57. Vickrey, W. *Testimony before the New York P.S.C., Case Number 26806*, Albany, New York, 1975.

58. Vickrey, W. "Some Implications of Marginal Cost Pricing for Public Utilities", *AER Papers and Proceedings*, May 1955.

59. Vickrey, W. "Responsive Pricing of Public Utility Services", *Bell Journal of Economics and Management Science*, Spring 1971.

60. Wenders, J.T. "Peak Load Pricing in the Electric Utility Industry" *Bell Journal of Economics*, Spring 1976.

61. Williamson, O.E. "Peak Load Pricing and Optimal Capacity Under Indivisibility Constraints" *AER*, September 1966.